

Pijera-Díaz, Héctor J.; Drachsler, Hendrik; Kirschner, Paul A.; Järvelä, S.

Profiling sympathetic arousal in a physics course: How active are students?

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:

formally and content revised edition of the original source in:

Journal of computer assisted learning 34 (2018) 4, S. 397-408, 10.1111/jcal.12271



Bitte verwenden Sie in der Quellenangabe folgende URN oder DOI /

Please use the following URN or DOI for reference:

urn:nbn:de:0111-dipfdocs-161603

10.25657/02:16160

<https://nbn-resolving.org/urn:nbn:de:0111-dipfdocs-161603>

<https://doi.org/10.25657/02:16160>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

DIPF | Leibniz-Institut für
Bildungsforschung und Bildungsinformation
Frankfurter Forschungsbibliothek
publikationen@dipf.de
www.dipfdocs.de

Mitglied der


Leibniz-Gemeinschaft

This is the peer reviewed version of the following article: „Profiling sympathetic arousal in a physics course: How active are students?“, which has been published in final form at

<https://onlinelibrary.wiley.com/doi/abs/10.1111/jcal.12271>.

This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Profiling sympathetic arousal in a physics course: how active are students?

Héctor J. Pijeira-Díaz^a, Hendrik Drachsler^{b,c,d}, Paul A. Kirschner^{a,d}, Sanna Järvelä^a

{Hector.PijeiraDiaz, Sanna.Jarvela}@oulu.fi

{Hendrik.Drachsler, Paul.Kirschner}@ou.nl

^a Faculty of Education, University of Oulu, Finland, P.O. BOX 2000, 90014

^b Goethe University Frankfurt, Germany

^c German Institute for International Educational Research (DIPF), Germany

^d Welten Institute, Open University of the Netherlands, Valkenburgerweg 177, 6401 AT

Heerlen

Abstract

Low arousal states (especially boredom) have been shown to be more deleterious to learning than high arousal states, though the latter have received much more attention (e.g. test anxiety, confusion and frustration). Aiming at profiling arousal in the classroom (how active students are) and examining how activation levels relate to achievement, we studied sympathetic arousal during two runs of an elective advanced physics course in a real classroom setting, including the course exam. Participants were high school students ($N = 24$) who were randomly selected from the course population. Arousal was indexed from electrodermal activity, measured unobtrusively via the Empatica E4 wristband. Low arousal was the level with highest incidence (60% of the lesson on average) and longest persistence, lasting on average three times longer than medium arousal and two times longer than high arousal level occurrences. During the course exam, arousal was positively and highly correlated ($r = .66$) with achievement as measured by the students' grades. Implications for a need to focus more on addressing low arousal states in learning are discussed, together with potential applications for biofeedback, teacher intervention and instructional design.

Keywords

learning activations, sympathetic arousal, electrodermal activity, biosensors, classroom, multimodal learning analytics

PROFILING SYMPATHETIC AROUSAL IN A PHYSICS COURSE: HOW ACTIVE ARE STUDENTS?

Students learn not by passively listening to teachers, but by actively expanding their own knowledge structures as they interpret and integrate new experiences, based on their prior knowledge (Scardamalia & Bereiter, 2014). Learning is a multifaceted, active process involving cognitive, affective and social aspects that leave a ‘physiological footprint’, namely, an increase in the bodily activation levels known as arousal, which can be measured. In general, arousal refers to the degree of physiological activation and responsiveness triggered by an event, object or situation, during people’s interaction with the environment (De Lecea, Carter, & Adamantidis, 2012; Juvina, Larue, & Hough, 2017). Also, seeking high arousal or avoiding boredom is a human goal (Boekaerts, Koning, & Vedder, 2006).

In learning, arousal is a particularly desirable goal, since it accounts for cognitive and/or affective activation of learners, which is a requirement for learning, as learning has been shown to be most effective when it is active. In his classic article, Hebb (1955) proposes that efficient learning is possible only in the waking, alert, responsive state, in which the level of arousal is high. He emphasizes ‘no arousal, no learning’. Empirical research shows that, for learning to be effective, it requires the active participation of the learner (Bjork, Dunlosky, & Kornell, 2013).

Despite what we know about arousal as a physiological concomitant of cognitive and affective processes, studying arousal and its physiological markers during learning is neither well-known nor extensive in the learning sciences, which have primarily utilized self-reports and trace methodologies (Winne & Perry, 2000). There are three essential reasons for not using physiological data: first, the technological equipment needed to collect it; second, a substantial

domain-specific knowledge base is needed to analyse and interpret it; and, third, the data from biosensors can be noisy and hard to relate to ‘traditional’ learning data (Palumbo et al., 2017).

There is no research available on how often students experience high arousal in a real classroom setting, how long different levels of activation last and to what extent different learning facets (e.g. task domain knowledge, mood, motivation, or group functioning) contribute to them. So far, researchers looking at the arousal dimension in learning processes have focused on the detection of students’ emotional states and the relation of their arousal levels to their achievement (Gillies et al., 2016). However, attempting to categorize the rich diversity of academic emotions in non-overlapping ways may result in insufficient discriminant validity (Pekrun, Goetz, Titz, & Perry, 2002). Therefore, the use of arousal as the activation dimension proposed in the circumplex model of affect (Russell, 1980) provides a more general and less compromised indicator which still captures an important dimension of affect (activation; see Figure 1). Moreover, arousal reflects cognitive processes, as well.

This study builds on the notion that learning is an active process involving cognitive and emotional processes, and that arousal is a physiological component of those processes. We look at learning activations from a physiological perspective. The approach chosen is necessarily exploratory due to the lack of studies characterizing arousal in the ecologically valid setting of an actual class. We profile sympathetic arousal in the classroom during two runs of an elective advanced high school physics course. Moreover, we examine its relation to performance on the course exam.

Next, we present the theoretical framework for the paper. We start by discussing the connection of arousal to cognitive and affective process, followed by the close link that such connection creates in turn between arousal and performance. We then situate arousal in the

naturalistic scenario for formal learning we are interested in, the classroom, by providing some related work and gaps identified in the literature of arousal in the classroom and during examinations. Since arousal is a complex response involving several physiological systems and operationalizations, we finalize our theoretical framework with the rationale behind using particularly the electrodermal system as the gateway to sympathetic arousal.

Cognition, affect and arousal

Arousal, the degree of physiological activation, originates from cognitive, affective and/or motor drivers (Critchley, 2002; Poh, Swenson, & Picard, 2010). In turn, changes in arousal inform cognitive processes (Critchley, Eccles, & Garfinkel, 2013), emotions (Schachter & Singer, 1962), and behaviours (Damasio, Tranel, & Damasio, 1991). The formation of students' self-efficacy beliefs relies partly on their arousal states (Bandura, 1982), while events of motivational significance elicit automatic changes in arousal (Critchley et al., 2013).

On the cognitive side, the notion of arousal has been related to cognitive processes like attention, memory and decision-making (Dawson, Schell, & Filion, 2017). Empirically, arousal has been shown to increase with attention in relation to eye-catching, engaging stimuli and attention-demanding tasks (Poh et al., 2010). Such a relationship might be explained by indications from clinical studies that a common neural substrate may be shared by arousal and attention mechanisms (Critchley, 2002). Emotionally arousing incidents and stimuli have been demonstrated to be much better remembered than those without emotional relevance (Marchewka et al., 2016). The representation of bodily changes in arousal may mediate the feeling of knowing ascribed to some familiar memories (Garfinkel, Critchley, & Pollatos, 2017). Arousal can influence memory encoding by modulating the selectivity of attention, and then

increasing attentional time on the arousing stimuli, thus, delaying disengagement (Fox, Russo, Bowles, & Dutton, 2001).

On the affective side, the classic cognition-arousal theory of emotion claims that emotional states are states of arousal whose labelling (e.g. boredom, anxiety, enjoyment and so on) is determined by the cognitive appraisal of the emotive stimulus (Schachter & Singer, 1962). The circumplex model of affect (Russell, 1980) characterizes emotion as bidimensional in an arousal-valence space (see Figure 1). The arousal axis (vertical) represents a continuum from deactivation to activation (bottom-up), while the valence axis (horizontal) represents a continuum from negative/unpleasant to positive/pleasant (left-right). As represented in Figure 1, building on such a model and particularizing for academic emotions, the four quadrants of the rectangular coordinate axes formed by arousal and valence have been matched to the quality of students' information processing (i.e. cognition), as well as to the type of strategies which they unfold (Pekrun et al., 2002). For example, positive activating emotions elicit flexible, creative learning strategies, such as elaboration, organization and critical evaluation, while negative activating emotions, on the other hand, may default to more rigid strategies, such as simple rehearsal and use of algorithmic procedures (Pekrun et al., 2002). Note that, in any case, the use of strategies is better for learning than the shallower, superficial processing of information characteristic of deactivating emotions, whether positive or negative.

Figure 1 about here

Boredom, a low arousal or deactivation state, has been shown to be more deleterious to learning than other cognitive-affective states that have received much more attention, such as anxiety, frustration and confusion (Baker, D'Mello, Rodrigo, & Graesser, 2010). As high arousal (i.e. energized) states, anxiety, frustration and confusion could be productive for learning if they

are properly regulated, and they might not even need remediation (Baker et al., 2010).

Conversely, boredom impairs learning by triggering behavioural or mental avoidance strategies when students face tasks which are either not engaging or which exceed their capabilities (Pekrun et al., 2002). Moreover, boredom is a more persistent state than other negative ones, whether activating or deactivating (Baker et al., 2010). Therefore, among the rich repertoire of academic emotions, there is a particularly special need to equip teachers with capabilities to detect bored students, especially in those acute cases of persistent or recurring boredom, so that interventions can be made to bring the student back onto the learning track. Boredom might be addressed, for example, by calibrating difficulty or providing more engaging tasks (Bosch et al., 2015).

The cognitive processes and affective states closely related to arousal here discussed, obviously have an impact on performance too, leading to a transitive relation between arousal and performance.

Arousal and performance

The relationship between arousal and performance—dictated by the so-called Yerkes-Dodson law (Yerkes & Dodson, 1908)—has been known for over a century. The Yerkes-Dodson law characterizes performance as a function of the arousal level following an inverted-U curve. When the arousal level is low, so is performance. As arousal rises due to increased attention, interest and engagement, performance increases up to a level of optimal arousal for performance at the top of the inverted-U shape. If arousal keeps rising to too high of a level because of excessive stress or strong anxiety, performance is impaired. Such impairment has been explained through a reduction of the processing capacity available for cognitive evaluation (Sanbonmatsu & Kardes, 1988). In the same vein, students with high levels of test anxiety perform worse on

tests, and their overall academic achievement is lower (Aritzeta et al., 2017). Test anxiety, located in the high arousal, negative valence area, has been shown as the academic emotion most often reported by the students, and can reduce working memory resources (Pekrun et al., 2002).

On some occasions, an external agent (e.g. a teacher) needs to give a team either feedback or additional information required to complete a task, while the task is being executed. Such an interruption may lead to a disruption of the work. Interruption-management systems based on arousal have been developed, whose results point to a significant increase in both task-performance and reported collaboration experience, when information was presented during arousal acceleration, compared to random interruption (Goyal & Fussell, 2017).

Studies of arousal in relation to performance, and to the cognitive processes and affective states behind performance have been carried out in a variety of settings. Next, we present some related work and gaps in the particular cases of the classroom and during examinations.

Arousal in the classroom

Though studies of arousal during classroom instruction are rare, some exceptions follow. Arroyo and colleagues (2009) used raw electrodermal activity (EDA) from high school students taking their regular mathematics class during five days, to determine their affective states in combination with self-reports, a camera for facial expression recognition, a pressure mouse and a posture analysis seat. More recently, Gillies and colleagues (2016) reported a case study on the in-class arousal of sixth graders during a unit of their science curriculum, in relation to their beliefs and engagement. They determined arousal through EDA by averaging the signal within time bins. Calderón (2016) tracked cardiovascular measures of arousal during two science education course presentations from two student teachers. However, to the best of our knowledge, arousal has not been measured throughout an entire course. Furthermore, studies

from some course sections which employed EDA to determine arousal have not used the frequency of the skin conductance responses to operationalize the level of arousal.

When it comes to ecologically valid learning situations, examinations have received much more attention than everyday classroom lessons, given the special interest on the relation between arousal and performance. Examinations are challenging situations often involving stress and anxiety, since they have not only psychological implications for the students, but also practical consequences for their progress in academic life, as well as their future. Real examinations may produce different responses than laboratory challenging situations because individuals evaluate examinations more seriously and are under more pressure to do well (Spangler, 1997). Learning scientists have long been interested in studying students' responses to exams (particularly test anxiety) and the implication they have for their performance (Pekrun et al., 2002). The physiological responses in examination situations have been studied from the cardiovascular system, the endocrine system and the immune system perspectives (Spangler, 1997). However, no examination studies have included pure measures of sympathetic activity (Spangler, 1997), as would be the case in EDA measures. Moreover, wristband EDA measures are less invasive and unobtrusive than the endocrine and immune system measurements involving saliva. Furthermore, the analysis of saliva to assess those physiological responses requires the use of medical laboratory equipment, while, for the wristband EDA measures, once the recording is done, the analysis is carried out using computer software. This study looks at the sympathetic arousal (as indexed by the electrodermal system next discussed) during a real exam and its relation to students' performance.

Electrodermal activity: the gateway to sympathetic arousal

As a complex response, arousal manifests through a variety of central (i.e. electrocortical), motor (i.e. muscle tension), autonomic (e.g. sweating, heart rate and blood pressure) and endocrine (i.e. hormonal) responses (American Psychiatric Association, 2000). In this study, we focus on sympathetic arousal, that is, arousal from the sympathetic branch of the autonomic nervous system, as measured through electrodermal activity.

Two types of arousal are distinguished based upon their measurement techniques: perceived arousal and physiological arousal (Reisenzein, 1983). Perceived arousal is measured through questionnaires, while physiological arousal is measured by means of biosensors. In particular, physiological arousal is determined through EDA (Edelberg, 1967), heart rate (Kreibig, 2010), electroencephalography (Hanoch & Vitouch, 2004) and pupilometry (Marchewka et al., 2016), among other techniques. For this study, we chose physiological measures of arousal since they are objective, in contrast to subjective self-reporting. From possible physiological measures, a number of features led to the selection of EDA, also known as galvanic skin response, although the former term is now preferred (Critchley, 2002; Dawson et al., 2017). First, the electrodermal system is the only one in the entire body solely innervated by the sympathetic nervous system (SNS). When it comes to arousal, the interest in the SNS, in contraposition to the parasympathetic nervous system, comes from its function of preparing the body for action, whether physical or intellectual. The SNS is the control system for the so-called fight-or-flight response, which reflects a state of activation. Conversely, the parasympathetic branch is responsible for bringing the body back to homeostasis, the state of equilibrium corresponding to the so-called rest-and-digest response, thus, counteracting arousal. In other words, EDA is a clean measure of sympathetic arousal, without parasympathetic interference.

Second, EDA has consolidated its place as a well-validated and widely accepted measure, and it is one of the simplest, most readily accessible indicators of arousal (Critchley, 2002; Neumann & Blanton, 1970). Extensive research has shown EDA to be one of the most popular bodily responses used in psychophysiological studies (Dawson et al., 2017; Kreibig, 2010).

EDA refers to the variations in the electrical properties of the skin as a result of perspiration. Basically, sweat secretion increases the skin conductivity since sweat is a good conductor (Poh et al., 2010). The changes are easily measurable with a properly spaced pair of electrodes. Perspiration results from either a thermoregulatory mechanism or a psychological process. Therefore, keeping temperature constant—23°C is recommended (Boucsein, 2012)—discards the thermoregulatory driver, and provides access to the psychological component.

The EDA signal is comprised of a slow-varying component known as the skin conductance level and a fast-changing component referred to as the skin conductance response (SCR) (Boucsein, 2012). SCRs, interpreted as arousal episodes, resemble peaks with a steep incline for the rising slope and an exponential-like decay (Benedek & Kaernbach, 2010b), as modelled in Figure 2.

Figure 2 about here

Common features to quantify arousal from the SCRs are their count, frequency, amplitude, and area under the curve (Benedek & Kaernbach, 2010b). The last two are more suitable to examine the arousal response to a particular stimulus of interest and are more sensitive to individual differences. Therefore, in this study we use count and frequency since they characterize the level of activation through number of arousal episodes, which is what we are interested in measuring naturalistically in the classroom.

Biosensors in the form of wearable devices, such as wristbands, increasingly facilitate the collection of physiological data (Schneider, Börner, van Rosmalen, & Specht, 2015), from where the arousal measure is then derived. The market availability, diversification and enhancement in technical specifications and adoption rates of these devices are on the rise (Swan, 2012). Moreover, their comfort of use and portability are attractive features for research as they allow taking measurements in the wild, outside of traditional laboratory environments. On the one hand, this leads to greater ecological validity, while, on the other hand, it affords online, objective and unobtrusive physiological measurements of learning processes. In such affordances of wearable sensors, we saw a way to reach our aims of an ecologically valid characterization of arousal in the classroom.

Aims

The aim of the study is twofold; first, to profile arousal in the classroom (How active are students?); second, to see how activation levels relate to performance on exams (i.e. academic achievement). The research questions are the following:

- RQ1. What is the incidence and persistence of arousal levels (low, medium and high) in the lessons of a science course?
- RQ2. How many students are simultaneously in the same arousal state (low, medium and high) during the lessons of a science course?
- RQ3. To what extent students' perceptions of their task domain knowledge, mood, motivation and group functioning explain in-class arousal?
- RQ4. How does arousal relate to students' academic achievement?

Method

Participants

The participants ($N = 24$) in the study were Finnish high school students, randomly selected from those enrolled in the regular, elective Advanced Physics course, which runs twice during the spring term. The number of participants in the study was determined by the Empatica[®] E4 wristband sensors (see *Materials*) available, which enabled us to track 12 students simultaneously during each course run. The gender distribution was six females (25%) and 18 males (75%), and their ages ranged from 16 to 17 years of age. They were high achievers in the preceding physics course, obtaining an average grade of 9.0 out of 10 ($SD = 0.6$).

Participation in the study was voluntary. An informed written consent was obtained from the students, who could revoke it at any time during the course. The study was approved by the relevant ethics committee.

Materials

The Empatica[®] E4 wristband (Empatica Inc., Cambridge, MA, U.S.A.) was used to record EDA continuously and unobtrusively during the course lessons and exam. While consumer-oriented wristband sensors available in the market, typically for fitness tracking, are of limited accuracy, the E4 is a research quality multi-sensor wristband (Garbarino, Lai, Tognetti, Picard, & Bender, 2014), and, therefore, significantly more expensive, which limited the number we had at our disposal (i.e. 12). The wristband embeds four sensors—EDA, photo-plethysmograph, thermometer and accelerometer—and has two mutually exclusive modes of operation: Bluetooth streaming and internal memory recording. The latter was chosen as more suitable for offline processing, given our exploratory intentions. The streaming mode, on the other hand, offers potential for real-time applications, such as biofeedback, learning dashboards or intervention

flags. The E4 wristband EDA sensor—with a sampling frequency of 4 Hz (i.e. four samples per second), uses the exosomatic method, which measures skin conductance in microSiemens (μS) by applying a small external current (Edelberg, 1967).

A dedicated online course was created via the Open edX (<https://open.edx.org/>) platform to aid teachers in guiding students throughout the lessons and to facilitate the students' review of the lessons in their self-study. The online course was designed according to the traditional course structure as provided by the teachers (e.g. course content, additional support learning material, simulations and videos). Open edX was chosen because of its customization affordances, as it is open source software. The students accessed Open edX on the iPads assigned to them by the school.

To assess student perception of cognitive, affective, motivational and collaborative aspects, a survey module was added at the end of each lesson's navigational menu on the online course. We chose single items given the limited time available for these measurements in the naturalistic setting of our study: a real high school course, packed with content and coursework, including both paper-and-pencil tasks and hands-on experiments. In this way, and since the survey was to be completed in every single lesson, we minimized the impact on the measurement environment (Winne & Perry, 2000). While multiple-item inventories are considered to be more reliable, rating scales based on single items may be economical and useful under specific circumstances (Pekrun et al., 2002). The design of the survey was inspired by the four single-item assessment devised by Schmitz and Skinner (1993) for an experiment in a similar setting, a naturalistic study with data collected in the classroom over four months. The resulting survey was used to obtain the students' ratings on their task domain knowledge, mood, motivation and group functioning for their lessons, based on a 1–10 scale (Table 1).

Table 1 about here

Procedure

The Advanced Physics course was selected to explore arousal in the classroom because it is elective and rated by the teachers as difficult, a combination that raises the potential number of SCRs in the EDA signal (i.e. sympathetic arousal). As it is an elective which the students themselves choose, we may assume that the students taking it are interested in the topic. Interest invites attention and engagement, both positively correlated to arousal. Once the students are engaged in their learning, difficulty increases cognitive and emotional arousal.

The course met for three 75 min lessons weekly. In all, the course consisted of 18 lessons, plus the exam. Topics of the course lessons included, amongst others, wave formation and interference, reflection and refraction, sound waves, speed and intensity of light. In general, the lessons began with a theoretical part (i.e. the teacher's explanation of the topic) followed by a collaborative practical part with paper-and-pencil problems or hands-on experiments and instructions and support material in the online environment. In the practical part, students put the learned physical theories, laws and principles to the test in groups of three. The groups remained the same for the duration of the course.

The classroom temperature was kept constant at 23°C to avoid thermoregulatory interference with the psychological EDA measurement (Boucsein, 2012). This did not require any intervention as it was the standard thermostat adjustment value in the school.

Students were instructed to wear the E4 wristband on their non-dominant hand in order to reduce the effect of movement artefacts as recommended in the literature, and to adjust the wristband strap neither loosely nor tightly so that the electrodes made proper but not excessive

contact with the skin, which would result in a pressure artefact (Edelberg, 1967). They put the wristband on at the beginning of each lesson and took it off at the end.

Students answered the survey in the Open edX platform at the end of each lesson.

Collected data

The data gathered for this exploratory study on classroom arousal involved EDA, the survey ratings, students' demographic information and grades (including those for the preceding physics course). On a nominal basis, 216,000 EDA samples were expected from each lesson (4 samples/second during 75 min from 12 devices; $4 \times 60 \times 75 \times 12$). However, some students were absent for some of the lessons. The attendance average was 10.9/12 ($SD = 1.3$) for the first run, and 10.5/12 ($SD = 1.3$) for the second. All students attended the exam. The total number of EDA samples collected was 6,426,492.

Analysis

The software Ledalab (version 3.4.9; <http://www.ledalab.de/>), based on MATLAB (The MathWorks, Inc., Natick, MA, U.S.A.), was used for EDA signal processing as recommended by the wristband manufacturer (Empatica Inc., 2015). The Ledalab algorithm assumes raw, unfiltered data as the input for decomposition analysis and feature extraction of the EDA signal (Benedek & Kaernbach, 2010a). Therefore, no signal pre-processing, such as cleaning or filtering, was performed on the EDA signal as obtained from the wristband sensor.

Traditionally, in research using EDA, SCRs (peaks) are detected using the classical trough-to-peak technique, from local minimum to local maximum (Boucsein, 2012; Edelberg, 1967). However, a subsequent SCR often happens during the decay of the previous one, leading to the superimposition of SCRs, which trough-to-peak is unable to detect separately, i.e. it is a biased measure. Several SCR decomposition approaches have been suggested to counteract this

effect, leading to the counting of individual SCRs, as such, and not as a superimposed group response. Currently, the recommended method for such decomposition is the continuous decomposition analysis (Benedek & Kaernbach, 2010a). This is the technique used here for the extraction of SCRs. In addition, one must decide on the amplitude threshold beyond which a change in conductance will be considered an SCR. Due to historical reasons—the smallest shift visible on paper chart recorders—the customary minimum amplitude used in EDA research is $0.05\ \mu\text{S}$ (Dawson et al., 2017). Although the advances in sensor technology—especially sensor resolution—currently enable lower thresholds down to $0.01\ \mu\text{S}$ (Braithwaite, Watson, Jones, & Rowe, 2013), for comparability purposes with previous studies, we chose to use the long-time standard of $0.05\ \mu\text{S}$.

The features extracted from the EDA signal were the SCR onset (after continuous decomposition analysis), and, subsequently, the SCR frequency on a minute basis. SCR frequency was calculated using a moving window approach, with a window width of one minute and a moving step of 250 ms—the sampling interval of the sensor—as parameters. In other words, the SCR frequency during the lesson was obtained for every time instant (resolution of 250 ms) after the first minute of the lesson, as it corresponded to the number of SCR onsets in the previous minute. Since no amplitude measure was considered, no correction in the form of standardization was needed for comparison across individuals.

The SCR frequency was used to categorise arousal into low, medium and high levels. Typically, a frequency of 1–3 peaks/min (ppm) occurs at rest (Dawson et al., 2017, p. 225), and, as frequency increases with the arousal level, values higher than 20 ppm are interpreted as high arousal (Boucsein, 2012, p. 222). Accordingly, frequencies of up to 3 ppm were labelled as low, from 20 ppm on were labelled as high, and anything in between was labelled as medium. For

illustration purposes, Figure 3 provides an example of a single arousal episode and a continuous high arousal interval of approximately 8 min, annotated in an excerpt of the SCR component of EDA extracted using Ledalab. The segment corresponds to one of the participants in one of the course lessons.

Figure 3 about here

Persistence was operationalized as the continuous duration of a certain arousal level (i.e. without switches of level in between). Persistence was computed for every single arousal level occurrence, with a resolution of 250 ms.

Arousal levels based on SCR frequency were used for RQ1 and RQ2, and total SCR count was used for RQ3 and RQ4, consistent with the target of the respective research questions.

Using the minimum of 1 ppm as the worst case, we discarded from the analysis all the EDA recording sessions where the total amount of SCR detected was less than 1 ppm times the lesson duration in min. In that way, we automatically excluded sessions where improper electrode-skin contact, due to wristband maladjustment, led to the measurement of basically noise rather than actual EDA signal. This devised technique served as a convenient aid to the manual visual inspection, recommended as the starting point for EDA analysis (Braithwaite et al., 2013). As a result, 137/420 (33%) of the EDA recording sessions were discarded as noisy. Additionally, 46/420 (11%) of the recording sessions were missing due to absent students, leaving 237/420 (56%) as valid EDA recording sessions. The previous numbers refer to the course lessons. Fortunately, there were no absent students during the exam, and no EDA recording sessions during the exam had to be discarded as noisy. From the 237 valid EDA recording sessions, 153 answered surveys were available, as, unfortunately, in some lessons,

students failed to complete the survey. Correlation analyses between the survey ratings and EDA-indexed arousal for RQ3 were then obviously performed on this subset of 153 cases.

Results

RQ1. Table 2 lists the descriptive statistics of the arousal levels' incidence, persistence and switches of level. Arousal levels were categorized into low, medium and high; exhaustive and mutually exclusive categories. Incidence was described in terms of percentage relative to the time of each lesson, which was 75 min. On average, low arousal was the clearly predominant state (60%) during the lessons. High arousal was the only level that did not occur for each student in every lesson (minimum of 0%). The minimum persistence for the three levels was common, and corresponded to the sampling interval of the EDA sensor (250 ms). Those were momentary changes oscillating around the threshold between two adjacent levels. On average, high arousal levels lasted for approximately 1 min (59 s). Low arousal was the most persistent state (mean = 151 s), reaching approximately 1 h (3632 s). There were at least six changes of state within the lessons, with an average of 48, representing a change every 1.5 min.

Table 2 about here

The distribution of the arousal levels' persistence is shown in Figure 4. The histograms are adjusted so that the vertical scale is the same, as it was comparable for the three levels. However, to improve the visualization, the horizontal axes are cut off at those points where the visibility vanishes, meaning very low frequency of persistence beyond those points. That is why the maximum persistence for the arousal levels, as shown in Table 2, is not visible in the figure. It is worth noting how these vanishing points differ for the three levels, with high arousal approximately doubling medium arousal, whereas low arousal, in turn, doubles high arousal.

Figure 4 about here

RQ2. Table 3 shows the descriptive statistics for the number of students in the same arousal state (i.e. low, medium or high arousal) during the lesson. The relatively low SD shows that the measure is quite stable. On average, the overwhelming majority of the students were in a low arousal state, while less than 25% of students were simultaneously in a level of activation, whether medium or high.

Table 3 about here

For a graphical perspective, Figure 5 shows the average number of students simultaneously in the same arousal state. The lines from top to bottom represent low, medium, and high arousal levels, respectively. The resolution of the chart is 250 ms, given by the EDA sensor sampling interval.

Figure 5 about here

RQ3. Descriptive statistics for the survey and the number of arousal episodes on a lesson basis are summarized in Table 4. According to the SD, the survey values were quite stable, but the number of arousal episodes varied significantly across lessons. The variables in Table 4 were tested for normality using the Shapiro-Wilk test. They were found to be non-normally distributed. Additionally, the number of arousal episodes per lesson were represented against the different survey item scores in a scatter plot to visually inspect for linearity. As nonlinearity was observed and the variables were not normally distributed, the nonparametric Spearman's correlation was used instead of Pearson's. Spearman's correlation determines the strength and direction of the monotonic relationship between two variables. Monotonicity was not observed in the scatter plot and accordingly it was reflected in low Spearman's correlation coefficients.

This is, no monotonic relationship was found between the number of arousal episodes per lesson and the students' self-reported task domain knowledge ($r_s = .17$), mood ($r_s = .14$), motivation ($r_s = .18$) or group functioning ($r_s = .21$).

Table 4 about here

RQ4. First, the normality and linearity assumptions of Pearson correlation were checked by means of the Shapiro-Wilk test ($p = .57$ for arousal; $p = .65$ for the grades) and a scatter plot respectively. Compliance with both assumptions was found, and so was a high, positive and statistically significant Pearson correlation ($r = .66$; $p = .02$) between the number of arousal episodes of the students during the course exam, and their achievement, as measured by their exam grades.

Discussion

In this exploratory study, we profiled sympathetic arousal in the classroom based on EDA measures taken continuously and unobtrusively during two runs of a high school advanced physics course. Our aim was motivated by the potential role of using arousal in learning research and practice, as a physiological concomitant of cognitive, affective, motivational and social activations.

RQ1 concerned the incidence and persistence of arousal levels (i.e. low, medium and high) within the course lessons. Low arousal emerged as the level of largest incidence, occurring, on average, during 60% of the lesson. This was a surprising result given the interest that can be presupposed in the students, since the course was elective, and interest, attention, engagement and arousal are closely connected. Yet, despite the assumed interest, students spent more than half of the class in a low activation state. In the worst cases, medium arousal occurred for 2 min and 15 s (3% minimum), while high arousal did not occur at all (0% minimum). As for the

persistence, on average, low arousal levels were sustained the longest (151 s), over three times longer than medium arousal levels and over two times longer than high arousal. Low arousal levels lasted continuously for up to over 1 h (3632 s maximum) in the worst case. That is a sustained, significant amount of the class that the student seems to be inactive and unproductive, since arousal is expected to rise with attention-demanding tasks (Critchley, 2002; Dawson et al., 2017; Poh et al., 2010), as it is reasonable to presuppose the coursework of an advanced physics course to be. This picture of pervasive low arousal in the classroom, in general, is aligned with a similar finding from a study involving three computer-based learning environments (Baker et al., 2010), where boredom (a low arousal state) was found to be the most persistent state, in contrast to others in the high arousal area, such as confusion, frustration and engaged concentration. The combination of the two studies (Baker et al.'s and this one) suggests that, in both digital and physical learning environments, low arousal levels are highly common. This might be considered, at best, to be neutral and, at worst, detrimental to learning. We do not mean that low arousal levels should not occur at all during learning, since, as much as moments of relaxation facilitate disengagement, they can also strengthen motivation for the next stage of learning (Pekrun et al., 2002), and they might serve recovery purposes necessary in between intense periods of cognitive efforts. Ideally, the proportion of low arousal should be small in comparison to states of activation (i.e. medium and high arousal).

RQ2 dealt with the simultaneity of arousal levels among students during the lessons. Closely related to the results for RQ1, on average, most of the students (9.3) were simultaneously in low arousal at any instant (resolution of 250 ms) during class, whereas fewer than three were active, whether in medium or high arousal. The worst case (i.e. the maximum) shows all students in low arousal. At any moment, only one student was in high arousal on average, reaching up to

a best case of half the students. The results provide a look from another angle at the predominance of low arousal discussed above, followed by medium and high arousal (see Figure 5).

RQ3 looked at the explanation of in-class arousal as coming from cognitive (task domain knowledge), affective (mood), motivational and collaborative (group functioning) aspects of learning in the classroom. However, the results show that none of these aspects explained more than 4% of the variance of arousal, which contradicts the theoretical expectations and findings from previous studies (Pekrun et al., 2002). Rather than challenging the connection of these learning aspects to arousal, our view is that this lack of correlation suggests that higher granularity measures of those constructs are needed (i.e. more frequent measures). The EDA signal to determine arousal had a granularity of 250 ms, while the survey granularity can be said to be 75 min, the duration of the lesson. To satisfy the data collection for the investigation, we would need to survey the students more often, but this would impose a higher load on the participants and might affect the study and the experimental setting more. In this study, we followed a minimum environmental intervention principle to maximize ecological validity in our quest to explore arousal as it occurs naturally in the classroom. However, these results suggest that, if we are to understand the extent to which cognitive, affective, motivational and collaborative aspects of learning impact in-class arousal, some ecological validity might have to be sacrificed so that more frequent measures are taken (e.g. using the experience sampling methodology, think aloud protocols or cross-fertilization from laboratory studies).

RQ4 regards the relationship of sympathetic arousal and performance in an examination situation (the course exam). The high positive correlation ($r = .66$) found indicates that there is a strong positive relationship between arousal during the exam and the grades the students

obtained; the more active the students, the higher the grades and vice versa. This result is aligned with the well-known Yerkes-Dodson law (Yerkes & Dodson, 1908), and places our observation space on the left half of the inverted-U curve that characterizes performance as a function of arousal according to the law. The inverted-U shape describes how performance improves with arousal, but only up to a point from where any further increase in arousal (e.g. excessive stress or strong anxiety) impairs performance. Since the participants were high achievers in the previous physics course, it is reasonable to presuppose that they would have high physics self-efficacy (Bandura, 1982), or confidence in their capacity to succeed in the exam. Therefore, stress and anxiety at the level that hinders performance was neither expected, nor observed. The Yerkes-Dodson law has been previously tested across a variety of cognitive tasks in laboratory settings (Sanbonmatsu & Kardes, 1988). Arousal during examinations has been explored in relation to academic emotional reactions (especially test anxiety), using a variety of physiological responses, namely, those from the cardiovascular, endocrine and immune systems, but not from the perspective of pure sympathetic arousal (Spangler, 1997). This study adds to the extant literature by providing evidence supporting the left half the Yerkes-Dodson curve (i.e. non-detrimental arousal) during an authentic exam situation, from a sympathetic arousal perspective as indexed by EDA.

Implications

The application of psychophysiology in human-computer interaction is a growing field with significant potential for smart personalized systems (Cowley et al., 2016). In learning, this translates to intelligent tutoring systems better able to support students towards successfully achieving their goals (Arroyo et al., 2009; Strauss et al., 2005). However, working in this field requires comprehension of an array of physiological signals and analysis techniques (Cowley et

al., 2016). By being aware of the learners' arousal levels from low to high, a tutor (human or computer-based) can adaptively change either the environment (e.g. challenge level of a problem or scenario) or otherwise interact with the learners to optimize their arousal levels for learning (Sottolare & Goodwin, 2017). Once the students' arousal levels are determined and accessible to teachers, a number of cognitive-behavioural modification interventions are available to adjust these levels in the best interest of learning such as a) stress inoculation therapy (focused on ways to direct and maintain attention and to modulate arousal), b) mental simulations (to control focus on the task or to take the student out of a state of boredom), and c) modifying the classroom environment (cf. Boekaerts & Corno, 2005).

Under closely controlled conditions of what is happening in the classroom at any time, it might be interesting to detect what exactly was the situation (e.g. task, active discussion, teacher's introduction of challenging concepts etc.) that brought the most students' arousal up to a high level. In real-time applications, signalling to the teachers those particular moments enables them to identify the activities which provoked most students to be active. This could have implications for instructional design and pedagogical practice and interventions, especially when complemented with other kinds of data. Figure 5 presents a general picture of the average number of students in the same arousal state using data from all the lessons of the two runs of the course. The potential of such a visualization for practice would be realized by providing the teachers with that information on a lesson basis, giving sort of a radiography of each lesson from the activation perspective. Teachers usually develop beliefs about what worked and what did not work when it comes to engaging students, based on their experiences and perceptions. Such beliefs that play a role in shaping their pedagogical practice could be either reinforced or adapted based on these objective data.

The clear predominance of low arousal (i.e. low activity or inactive students) in class, even when students themselves chose to enrol in the course, which presupposes interest, might support the view that, too often, school tasks and practices are inappropriate for the intended learning goals (Scardamalia & Bereiter, 2014), thus failing to engage even interested learners. Moreover, it supports the claim that more attention should be paid and resources dedicated to tackling the problem of bored students (or inactive in general), than dealing with confused or frustrated students (Baker et al., 2010), since the latter both happen less often and have shorter persistence on average.

This exploratory study serves as a proof-of-concept for biosensor-supported learning dashboard applications, either in real-time or retrospectively. The field of learning dashboard itself has had an exploratory nature so far, but is slowly considering more evaluations of the effects of these tools in real courses (for a review, see Schwendimann et al., 2016). However, a recent literature study analysed the extent to which theories and models from learning sciences have been integrated into the development of learning dashboards aimed at learners. The analysis revealed that very few dashboard evaluations take into account the educational concepts which were used as a theoretical foundation for their design (Jivet, Scheffel, Drachsler, & Specht, 2017, 2018). The metrics of in-class arousal considered in this study could be incorporated into a new kind of ‘biosensor learning dashboard’ for students to be aware of how active they are and stimulate reflection on how they could increase engagement; moreover, this could enable teachers to monitor the number of active students at any time during class, and to realize how these activation levels relate to the different tasks they plan for their instruction. Furthermore, as biosensor learning dashboards can bridge the gap between learning sciences and data analytics, they are also a strong means for educational researchers. Researchers could test different

educational concepts and theories, e.g. on activation from a physiological perspective, and therefore enrich the observations of earlier studies that have been based purely on subjective measures.

Exploring the physiological dimension of learning, in addition to the applications it has in its own right, has a potential for stronger research in which this data modality is triangulated with traditional measures, such as surveys and digital traces (Järvelä, Malmberg, Haataja, Sobocinski, & Kirschner, 2018, submitted). Triangulation approaches have become a widespread recommendation to avoid relying exclusively on self-reporting (Boekaerts & Corno, 2005).

On the methodological side, building on the extant psychophysiological literature, a simple yet innovative technique was presented to easily discard noisy EDA recordings, based on the minimum of 1 ppm. This automatic procedure allows for integration of this technique in future applications for learning dashboards and/or intelligent tutoring systems as discussed above.

Limitations

The present study takes a step towards a more holistic research on learning by exploring a non-traditional data modality in the field, as is EDA. However, the study is not without its limitations. The number of participants ($N = 24$) was limited by the available E4 wristbands, which are quite expensive compared to the consumer-oriented devices, since they are specifically designed for research purposes and of better quality. Nonetheless, 44% of the nominal expected EDA recorded sessions were unavailable for analysis due to either absent students or not of a good enough quality due to the limitations of the current technology. This is a similar amount to that reported by Arroyo and colleagues (2009) from a study also using biosensors in a science high school classroom. Further technological developments are needed to systematically maximize

the amount of valid data. One such approach would be to use the wristband in streaming mode to send the data to a server, which in real time would assess the quality of the signal on a minute basis and raise an alert for readjustment in case of poor quality. A first approach to such an infrastructure has been made by Di Mitri et al. (2017).

Future work

Individuals vary in their ability to perceive their autonomic physiological state, known as interoceptive awareness, which they need to inform cognitions and behaviours (Critchley et al., 2013). For example, people rely partly on information from their physiological state in judging their capabilities (Bandura, 1982). An experiment design where the physiological measures and features used here (and many others still to be explored since there is a vast number of physiological features) are fed back to the students in real time would allow for studying the impact of this type of information on learning. How useful would it be for students to be supported by objective physiological features for their metacognitive skills in learning?

The present study has characterized arousal by frequency of EDA responses. Future work should also consider amplitude measures of arousal (e.g. SCR amplitudes and area under the curve). Experiment designs following the stimulus-response paradigm are needed to measure the area under the EDA response evoked by a stimulus within the response window.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Aritzeta, A., Soroa, G., Balluerka, N., Muela, A., Gorostiaga, A., & Aliri, J. (2017). Reducing anxiety and improving academic performance through a biofeedback relaxation training program. *Applied Psychophysiology and Biofeedback*, 42, 193–202.
<https://doi.org/10.1007/s10484-017-9367-z>
- Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. In *Proceedings of the Conference on Artificial Intelligence in Education* (pp. 17–24). Amsterdam: IOS Press.
- Baker, R. S., D’Mello, S. K., Rodrigo, M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: the incidence and persistence of affect during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68, 223–241. <https://doi.org/10.1016/j.ijhcs.2009.12.003>
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122–147. <https://doi.org/10.1037/0003-066X.37.2.122>
- Benedek, M., & Kaernbach, C. (2010a). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190, 80–91.
<https://doi.org/10.1016/j.jneumeth.2010.04.028>
- Benedek, M., & Kaernbach, C. (2010b). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47, 647–658. <https://doi.org/10.1111/j.1469-8986.2009.00972.x>
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: beliefs, techniques,

- and illusions. *Annual Review of Psychology*, 64, 417–444. <https://doi.org/10.1146/annurev-psych-113011-143823>
- Boekaerts, M., & Corno, L. (2005). Self regulation in the classroom: a perspective on assessment and intervention. *Applied Psychology*, 54, 199–231. <https://doi.org/10.1111/j.1464-0597.2005.00205.x>
- Boekaerts, M., Koning, E. de, & Vedder, P. (2006). Goal-directed behavior and contextual factors in the classroom: an innovative approach to the study of multiple goals. *Educational Psychologist*, 41, 33–51. https://doi.org/10.1207/s15326985ep4101_5
- Bosch, N., D'Mello, S., Hall, F., Baker, R. S., Shute, V., & Wang, L. (2015). Automatic detection of learning-centered affective states in the wild. In *Proceedings of the 2015 International Conference on Intelligent User Interfaces (IUI 2015)*.
- Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). New York, US: Springer.
- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, 49, 1017–1034.
- Calderón, O. (2016). Oximetry: a reflective tool for the detection of physiological expression of emotions in a science education classroom. *Cultural Studies of Science Education*, 1–15. <https://doi.org/10.1007/s11422-016-9731-y>
- Cowley, B., Filetti, M., Lukander, K., Torniainen, J., Henelius, A., Ahonen, L., ... Jacucci, G. (2016). The psychophysiology primer: a guide to methods and a broad review with a focus on human-computer interaction. *Foundations and Trends in Human-Computer Interaction*, 9, 150–307. <https://doi.org/10.1561/11000000065>
- Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *Neuroscientist*, 8,

132–142. <https://doi.org/10.1177/107385840200800209>

Critchley, H. D., Eccles, J., & Garfinkel, S. N. (2013). Interaction between cognition, emotion, and the autonomic nervous system. In R.M. Buijs & D. F. Swaab (Eds.), *Handbook of Clinical Neurology* (Vol. 117, pp. 59–77). Elsevier B.V. <https://doi.org/10.1016/B978-0-444-53491-0.00006-7>

Damasio, A. R., Tranel, D., & Damasio, H. C. (1991). Somatic markers and the guidance of behavior: theory and preliminary testing. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal Lobe Function and Dysfunction*. New York, US: Oxford University Press.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 217–243). Cambridge, UK: Cambridge University Press.

De Lecea, L., Carter, M. E., & Adamantidis, A. (2012). Shining light on wakefulness and arousal. *Biological Psychiatry*, *71*, 1046–1052.
<https://doi.org/10.1016/j.biopsych.2012.01.032>

Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., & Specht, M. (2017). Learning Pulse: a machine learning approach for predicting performance in self-regulated learning using multimodal data. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 188–197). Vancouver: ACM.
<https://doi.org/10.1145/3027385.3027447>

Edelberg, R. (1967). Electrical properties of the skin. In C. C. Brown (Ed.), *Methods in psychophysiology* (pp. 1–53). Baltimore, US: Williams & Wilkins.

Empatica Inc. (2015). Recommended tools for signal processing and data analysis. Retrieved March 1, 2018, from <https://support.empatica.com/hc/en-us/articles/202872739->

Recommended-tools-for-signal-processing-and-data-analysis

- Garbarino, M., Lai, M., Tognetti, S., Picard, R. W., & Bender, D. (2014). Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on* (pp. 39–42). Athens, Greece: IEEE.
<https://doi.org/10.4108/icst.mobihealth.2014.257418>
- Garfinkel, S. N., Critchley, H. D., & Pollatos, O. (2017). The interoceptive system: implications for cognition, emotion and health. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 427–443). Cambridge, UK: Cambridge University Press.
- Gillies, R. M., Carroll, A., Cunningham, R., Rafter, M., Palghat, K., Bednark, J., & Bourgeois, A. (2016). Multimodal representations during an inquiry problem-solving activity in a year 6 science class: a case study investigating cooperation, physiological arousal and belief states. *Australian Journal of Education*, 60, 111–127. <https://doi.org/10.1177/0004944116650701>
- Goyal, N., & Fussell, S. R. (2017). Intelligent interruption management using electro dermal activity based physiological sensor for collaborative sensemaking. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. ACM.
<https://doi.org/10.1145/3130917>
- Hanoch, Y., & Vitouch, O. (2004). When less is more: information, emotional arousal and the ecological reframing of the Yerkes–Dodson law. *Theory & Psychology*, 14, 427–452.
<https://doi.org/10.1177/0959354304044918>
- Hebb, D. O. (1955). Drives and the CNS (conceptual nervous system). *Psychological Review*, 62, 243–254. <https://doi.org/10.1037/h0041823>

- Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: pitfalls of learning analytics dashboards in the educational practice. In L. É., H. Drachsler, V. K. B. J., & P.-S. Mar (Eds.), *Data Driven Approaches in Digital Education. EC-TEL 2017. Lecture Notes in Computer Science* (Vol. 10474, pp. 82–96). Springer, Cham.
- Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2018). License to evaluate: preparing learning analytics dashboards for educational practice. In *Proceedings of International Conference on Learning Analytics and Knowledge*. Sydney: ACM.
<https://doi.org/10.1145/3170358.3170421>
- Juvina, I., Larue, O., & Hough, A. (2017). Modeling valuation and core affect in a cognitive architecture: The impact of valence and arousal on memory and decision-making. *Cognitive Systems Research*, 48, 4–24. <https://doi.org/10.1016/j.cogsys.2017.06.002>
- Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2018). What multimodal data can tell us about the students' regulation of their learning process?
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biological Psychology*, 84, 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Marchewka, A., Wypych, M., Moslehi, A., Riegel, M., Michalowski, J. M., & Jednoróg, K. (2016). Arousal rather than basic emotions influence long-term recognition memory in humans. *Frontiers in Behavioral Neuroscience*. <https://doi.org/10.3389/fnbeh.2016.00198>
- Neumann, E., & Blanton, R. (1970). The early history of electrodermal research. *Psychophysiology*, 6, 453–475. <https://doi.org/10.1111/j.1469-8986.1970.tb01755.x>
- Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., & Goodwin, M. S. (2017). Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review*, 21, 99–141.

<https://doi.org/10.1177/1088868316628405>

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: a program of qualitative and quantitative research.

Educational Psychologist, 37, 91–105. <https://doi.org/10.1207/S15326985EP3702>

Poh, M. Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assesment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*,

57, 1243–1252. <https://doi.org/10.1109/tbme.2009.2038487>

Reisenzein, R. (1983). The Schachter theory of emotion: two decades later. *Psychological*

Bulletin, 94, 239–264. <https://doi.org/10.1037/0033-2909.94.2.239>

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social*

Psychology. <https://doi.org/10.1037/h0077714>

Sanbonmatsu, D. M., & Kardes, F. R. (1988). The effects of physiological arousal on

information processing and persuasion. *Journal of Consumer Research*, 15, 379–385.

<https://doi.org/10.1086/209175>

Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: theory,

pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 397–417). New York: Cambridge University Press.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of

emotional state. *Psychological Review*, 69, 379–399. <https://doi.org/10.1037/h0046234>

Schmitz, B., & Skinner, E. (1993). Perceived control, effort, and academic performance:

Interindividual, intraindividual, and multivariate time-series analyses. *Journal of*

Personality and Social Psychology, 64, 1010–1028. [https://doi.org/10.1037//0022-](https://doi.org/10.1037//0022-3514.64.6.1010)

3514.64.6.1010

- Schneider, J., Börner, D., van Rosmalen, P., & Specht, M. (2015). Augmenting the senses: a review on sensor-based learning support. *Sensors*, *15*, 4097–4133.
<https://doi.org/10.3390/s150204097>
- Schwendimann, B. A., Rodriguez-Triana, M., Vozniuk, A., Prieto, L., Boroujeni, M., Holzer, A., ... Dillenbourg, P. (2016). Perceiving learning at a glance: a systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*.
<https://doi.org/10.1109/TLT.2016.2599522>
- Sottolare, R. A., & Goodwin, G. A. (2017). Adaptive instructional methods to accelerate learning and enhance learning capacity. In A. Bruzzone, E. Cayirci, & R. Sottolare (Eds.), *Proceedings of the International Defense and Homeland Security Simulation Workshop* (pp. 42–50). Barcelona, Spain: CAL-TEK SRL.
- Spangler, G. (1997). Psychological and physiological responses during an exam and their relation to personality characteristics. *Psychoneuroendocrinology*, *22*, 423–441.
[https://doi.org/10.1016/S0306-4530\(97\)00040-1](https://doi.org/10.1016/S0306-4530(97)00040-1)
- Strauss, M., Reynolds, C., Hughes, S., Park, K., Mcdarby, G., & Picard, R. W. (2005). The HandWave Bluetooth skin conductance sensor. In *International Conference on Affective Computing and Intelligent Interaction* (pp. 699–706). Berlin: Springer.
- Swan, M. (2012). Sensor mania! The internet of things, wearable computing, objective metrics, and the quantified self 2.0. *Journal of Sensor and Actuator Networks*, *1*, 217–253.
<https://doi.org/10.3390/jsan1030217>
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego, CA, US: Academic Press. <https://doi.org/10.1016/B978-012109890-2/50045-7>

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459–482.

<https://doi.org/10.1037/h0073415>

Tables

Table 1

Lesson-wise survey of cognitive, affective, motivational, and collaborative facets

| Facet | Question |
|-----------------------|-----------------------------------------------------------------------------------------------------------------|
| Task domain knowledge | On a scale of 1 (you knew nothing) to 10 (you knew everything), to what extent did you know what to do? |
| Mood | On a scale of 1 (very bad) to 10 (very good), how did you feel? |
| Motivation | On a scale of 1 (absolutely unmotivated) to 10 (fully motivated), how motivated were you to do the class tasks? |
| Group functioning | On a scale of 1 (very bad) to 10 (very well), how did your group work? |

Table 2

Incidence, persistence, and changes of arousal levels

| Arousal levels | | Min | Max | Mean | SD |
|----------------------------------|----------------|--------|--------|-------|-------|
| Incidence (% of lesson duration) | Low arousal | 10% | 95% | 60% | 24% |
| | Medium arousal | 3% | 65% | 24% | 11% |
| | High arousal | 0% | 74% | 17% | 19% |
| Persistence (s) | Low arousal | 0.25 s | 3632 s | 151 s | 297 s |
| | Medium arousal | 0.25 s | 2204 s | 45 s | 57 s |
| | High arousal | 0.25 s | 1569 s | 59 s | 101 s |
| Changes of level | Count | 6 | 128 | 48 | 25 |

Table 3

Descriptive statistics of the number of students simultaneously in the same arousal state during the lessons

| Number of students simultaneously in specific states of arousal | Min | Max | Mean | SD |
|-----------------------------------------------------------------|-----|-----|------|-----|
| Low arousal | 3 | 12 | 9.3 | 1.7 |
| Medium arousal | 0 | 8 | 1.6 | 1.3 |
| High arousal | 0 | 6 | 1.1 | 1.2 |

Table 4

Descriptive statistics for the survey and the total arousal episodes on a lesson basis

| | Min | Max | Mean | SD |
|-------------------------------|-----|------|-------|-------|
| Task knowledge | 4 | 10 | 9.0 | 1.1 |
| Mood | 2 | 10 | 8.8 | 1.2 |
| Motivation | 4 | 10 | 9.0 | 1.1 |
| Group functioning | 3 | 10 | 9.3 | 1.1 |
| Total arousal episodes/lesson | 79 | 2239 | 601.2 | 528.8 |

Figures

Figure 1. Cognitive-motivational model on the effects of emotions, characterized by arousal and valence. Based on Pekrun et al., 2002.

Figure 2. Mathematical model of a skin conductance response, tantamount to an arousal episode.

Figure 3. Example of single arousal episode and high arousal interval in an excerpt of skin conductance response signal extracted using Ledalab from one student in one lesson.

Figure 4. Distribution of persistence in s for the three arousal levels: low, medium and high.

Figure 5. Average number of students simultaneously in the same arousal state during the lessons (top-down: in low, in medium and in high arousal).

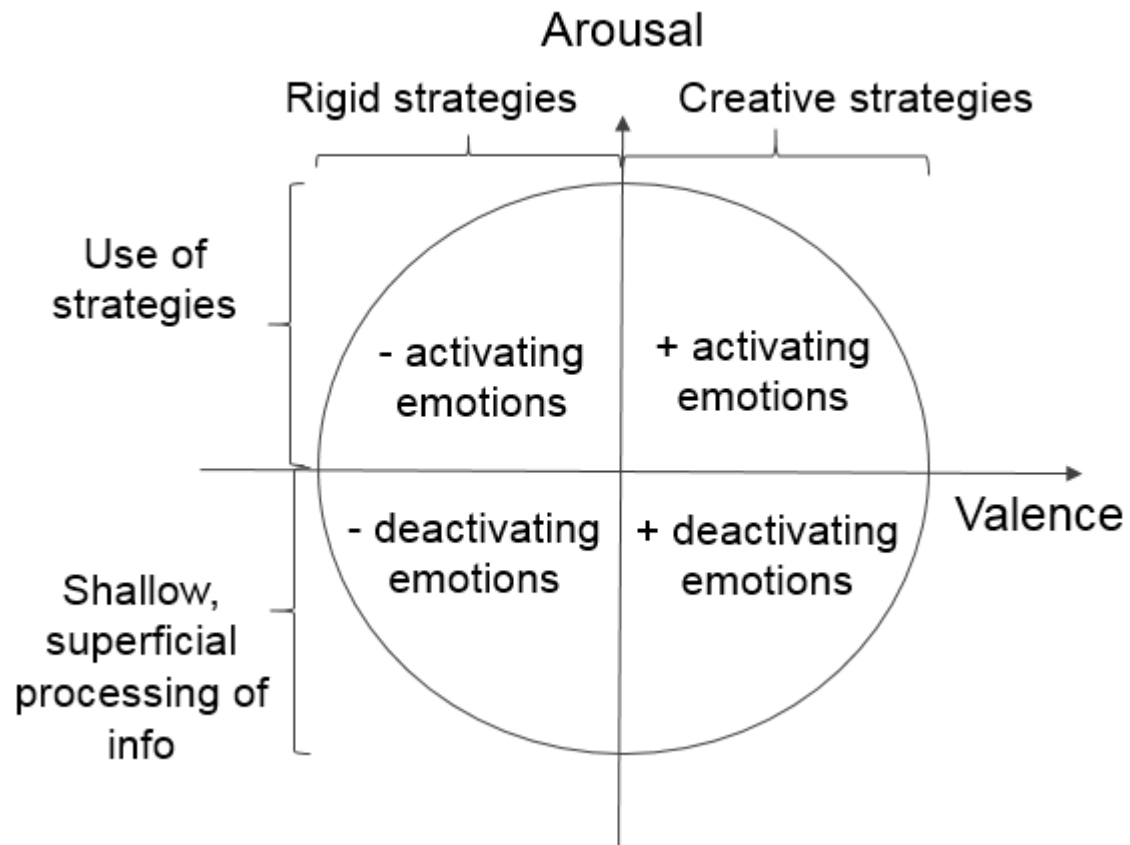


Figure 1

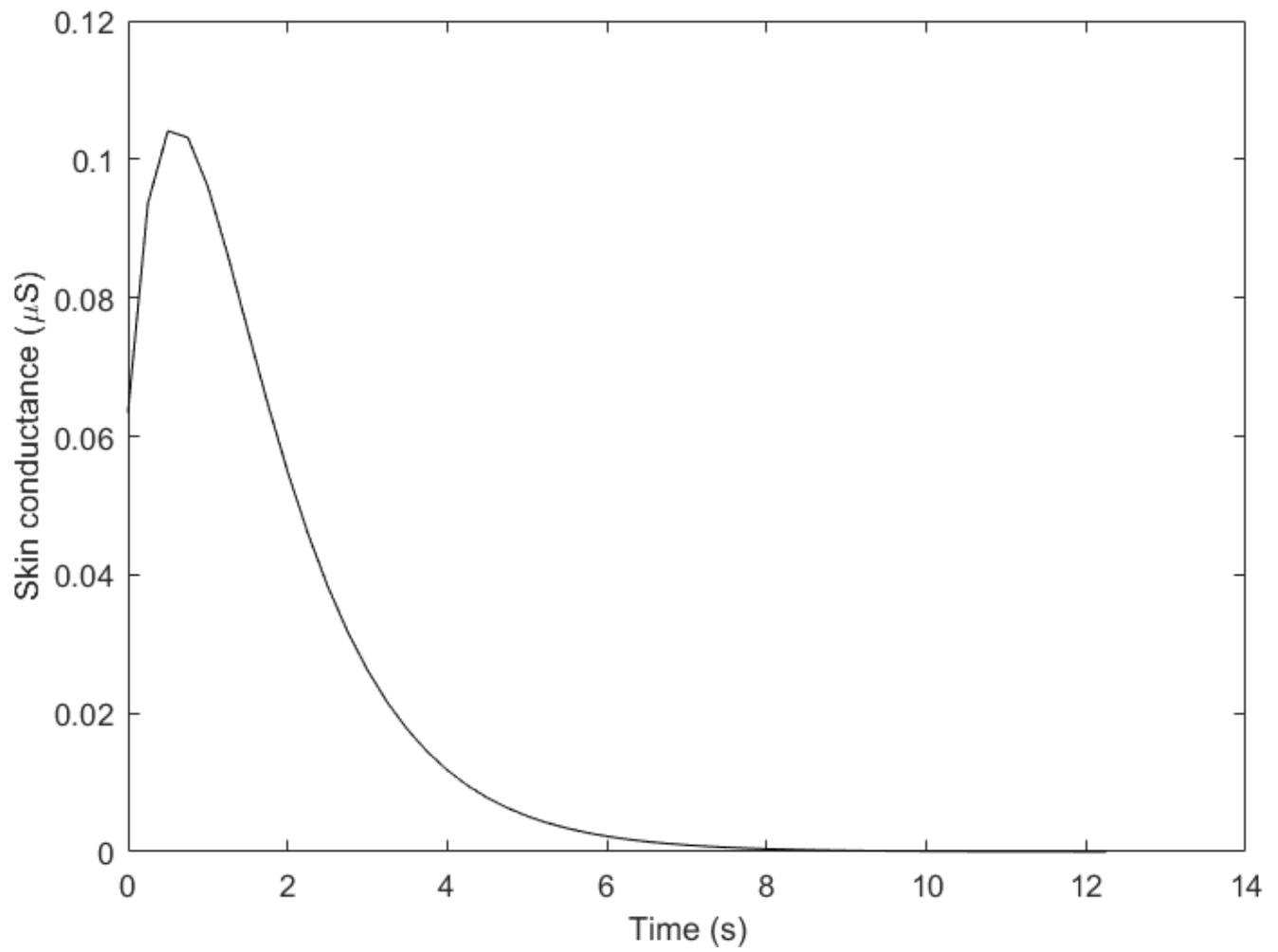


Figure 2

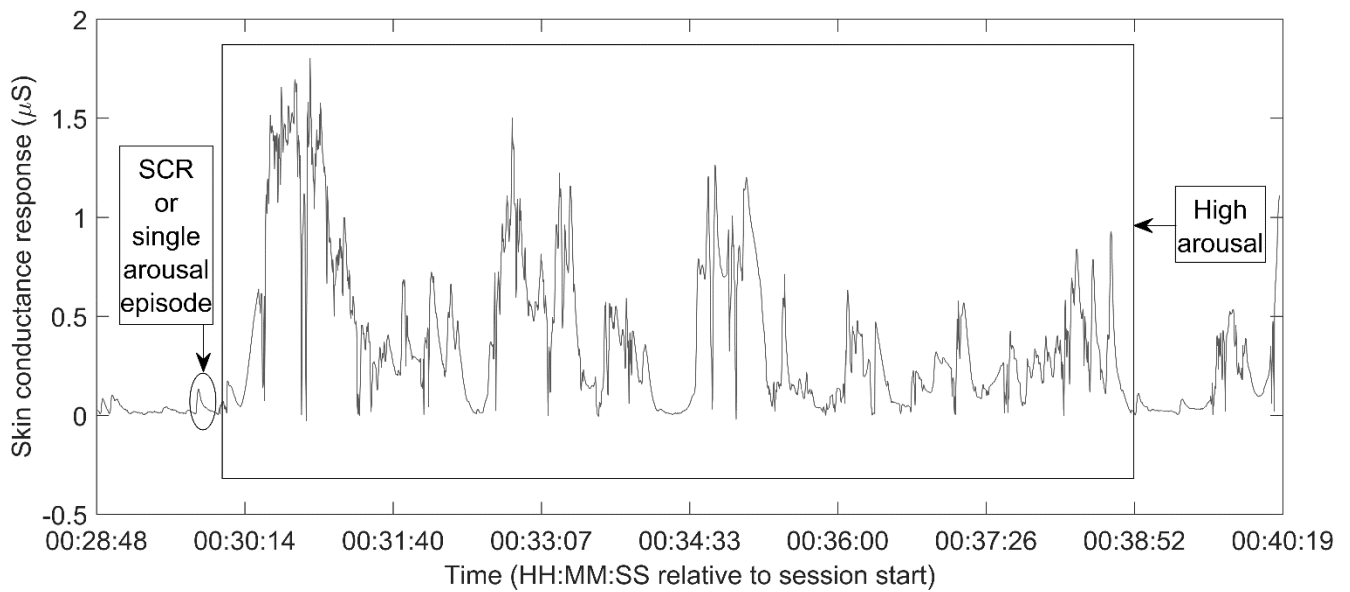


Figure 3

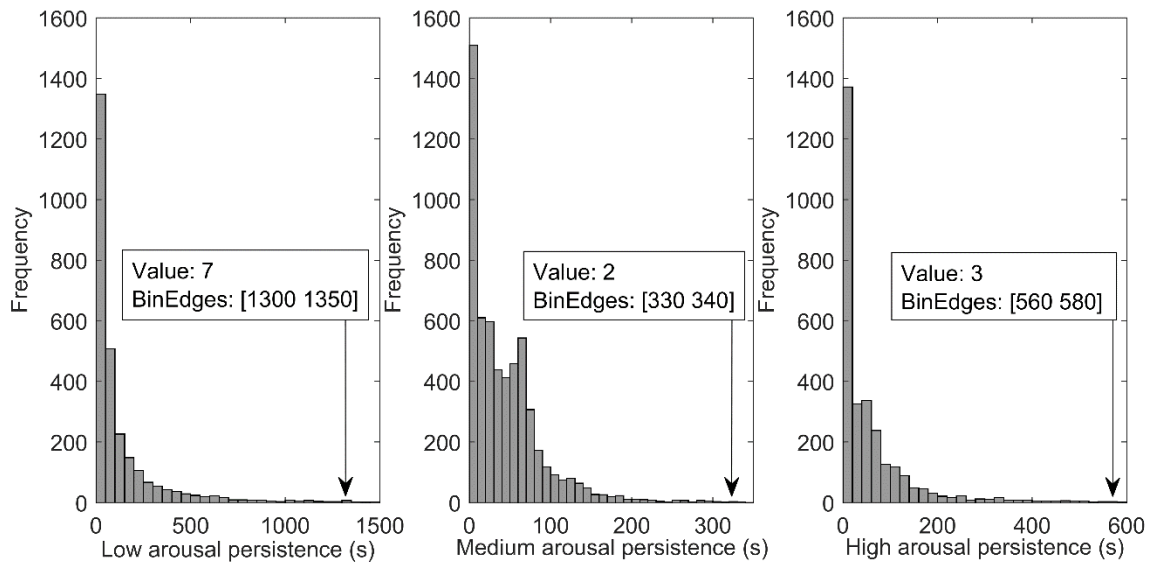


Figure 4

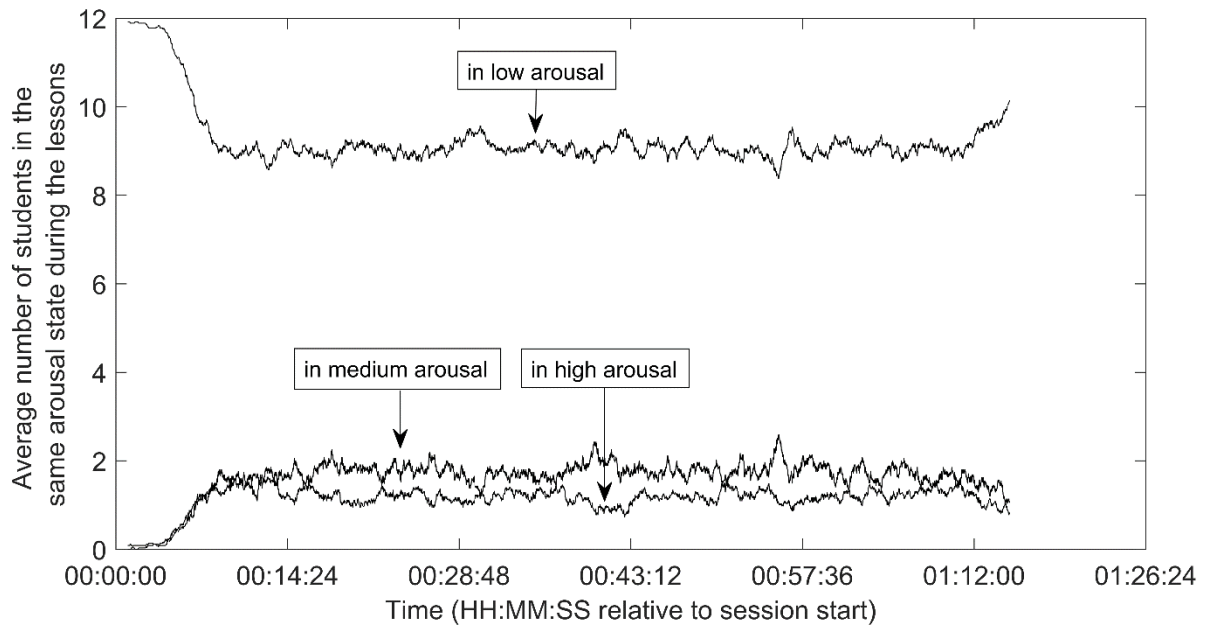


Figure 5

Acknowledgements

This study is funded by the Academy of Finland, project number 275440 (SLAM). The experiment could not have been carried out without the collaboration of teachers, high school students, and administrative personnel from the Oulu University Teacher Training School (Norssi).